AFHRL-TR-70-9

AIR FORCE

HUMAN RESOURCES

LABORATORY

PROFICIENCY DIFFERENCES OF PILOT AND NAVIGATOR F-4 SECOND-SEAT CREWMEMBERS: A SOUTHEAST ASIA EVALUATION

By

C. Wayne Shore, Capt, USAF
Charles R. Curran, Capt, USAF
Forrest R. Ratliff, LtCol, USAF
John R. Chiorini, Capt, USAF

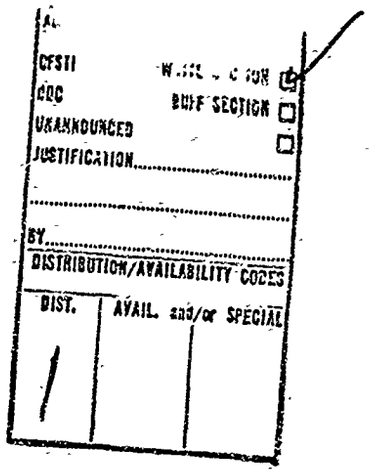
PERSONNEL RESEARCH DIVISION
Lackland Air Force Base, Texas

April 1970

AIR FORCE SYSTEMS COMMAND

BROOKS AIR FORCE BASE, TEXAS

33



NOTICE

Composition: Hope De La Cruz
Mary L. Alvarado
C&MS 3469, 3471, 3497, 3501, 3560

AFHRL-TR-70-8

April 1970

PROFICIENCY DIFFERENCES OF PILOT AND NAVIGATOR F-4 SECOND-SEAT CREWMEMBERS: A SOUTHEAST ASIA EVALUATION

By

C. Wayne Shore, Capt, USAF
Charles R. Curran, Capt, USAF
Forrest R. Ratliff, LtCol, USAF
John R. Chiorini, Capt, USAF

PERSONNEL RESEARCH DIVISION
AIR FORCE HUMAN RESOURCES LABORATORY
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas

FOREWORD

This report describes the development and use of a rating scale to measure F-4 crew performance differences under combat conditions. The research was carried out under Project 6323, Personnel Management Research and Development; Task 632305, Development of Statistical and Mathematical Procedures to Facilitate Personnel Research. The findings of this report were made in partial response to a Requirement for Personnel Research, RPR 68-16, originated by APGC (PGOT).

This report has been reviewed and is approved.

John G. Dailey, Colonel, USAF
Commander

ABSTRACT

A proficiency rating form was administered to F-4 crewmembers following each of a series of combat missions. The rating form was one of several data collection instruments developed as part of Project Combat Team to gather behavioral data designed to relate the differences in the training of pilots and navigators to operational performance. The rating categories were designed to measure proficiency in ten second-seater functions and three general characteristics related to second-seater proficiency. Using an 11-point rating scale, aircraft commanders compared their second-seat crewmember with second-seaters of equal combat experience. Differences between pilot and navigator second-seaters on each rating category were tested by a multiple linear regression analysis Data from other instruments relevant to evaluation of mission success were also presented. Several conclusions were drawn on the basis of these data: (*a*) It is possible to obtain reliable, valid proficiency data in a combat environment. (*b*) Proficiency ratings can be used to measure the performance differences between two groups in an operational setting. (*c*) By the 30th mission neither group of second-seaters demonstrated a general superiority in performance over the other. (Differences between pilots and navigators early in their combat experience were sharply reduced by mission segment 26 to 30, except in performance of the task Understanding Radio Communications.) (*d*) The early proficiency differences between pilot and navigator second-seaters appeared to be a function of training differences. Therefore, it may be possible to reduce or eliminate these differences through modifications in crew training programs. (*e*) During the first 30 combat missions, pilots flew on more missions rated as completely successful than did navigators. Although the differences were small, they were statistically significant.

SUMMARY

Shore, C.W., Curran, C.R., Ratliff, F.R., and Chiorini, J.R. *Proficiency differences of pilot and navigator F-4 second-seat crewmembers: A Southeast Asia evaluation.* AFHRL-TR-70-9. Lackland AFB, Tex.: Personnel Research Division, Air Force Human Resources Laboratory, April 1970.

Problem

At the request of the Department of Defense, the Air Force initiated a study to determine the feasibility of replacing pilots with navigators as second-seat crewmembers in F-4 aircraft. The Department of Defense estimated the change could save the Air Force $400 million in training costs over a 5-year period. The Air Force felt that the change in F-4 crew composition would be feasible only if it could be demonstrated that the use of navigators in the second seat would not have a detrimental impact on combat effectiveness or aircraft safety. Project Combat Team was established to develop and conduct a series of tests to determine if there were significant differences in the combat performance of navigator and pilot F-4 second-seaters. Included in the plan was a comparison of the combat performance of crews containing a pilot or a navigator in the second seat.

Approach

Data collection instruments, including a proficiency rating scale, were developed by personnel measurement psychologists following in-depth interviews with F-4 instructor pilots. The purpose of the proficiency rating scale was to measure differences between pilot and navigator second-seaters in the proficiency with which they perform certain critical inflight functions during combat missions.

The proficiency rating scale was administered at two Southeast Asia air bases where equal numbers of newly trained pilot and navigator second-seaters were assigned. In making ratings, aircraft commanders compared their pilot or navigator crewmember with F-4 second-seaters of equal combat experience. Information was also gathered from intelligence and crew debriefings concerning the degree to which mission objectives were achieved. During the data collection period, combat missions were limited to South Vietnam and the southern portion of North Vietnam.

Results

There was no general superiority in performance of one group of second-seaters (pilots or navigators) over the other across the tasks rated. Early in their combat experience, navigators were rated as more proficient in the Use of Radar, while pilots were rated as more proficient in Understanding Radio Communications and Visual Target Acquisition. However, by the time the second-seaters accumulated the experience of from 26 to 30 combat missions, these differences were sharply reduced, except for the performance of the task Understanding Radio Communications, in which case pilots retained their superiority. There were no instances recorded in over fourteen hundred missions where either a pilot or navigator second-seater endangered the safety of the aircraft. Pilot second-seaters had a higher percentage of missions evaluated as successful than did navigators. The difference was small but statistically significant. Further research is required to determine whether the differences in mission success would also be diminished by increased combat experience.

Conclusions

1. Performance characteristics related to crewmember proficiency can be identified and reliably measured in a combat environment.

2. There was no general superiority of performance of either pilot or navigator second-seat crewmembers, although pilots were consistently better at Understanding Radio Communications.

3. The decrease in the magnitude of proficiency differences which occurred between pilot and navigator second-seaters early in their combat experience indicates that the differences are a function of differences in the training of pilots and navigators. It may be possible, therefore, to reduce or eliminate these differences through modifications in crew training programs.

This summary was prepared by C.R. Curran, Armed Forces Radiobiology Research Institute, Defense Atomic Support Agency, Bethesda, Maryland.

TABLE OF CONTENTS

LIST OF TABLES

PROFICIENCY DIFFERENCES OF PILOT AND NAVIGATOR F-4 SECOND-SEAT CREWMEMBERS: A SOUTHEAST ASIA EVALUATION

I. PROBLEM

Background

In mid 1967, the Department of Defense requested that the Air Force study the feasibility of replacing pilot rear-seat crewmembers in F-4 aircraft with navigators. The present crew configuration of the F-4 places an aircraft commander in the front cockpit and another rated pilot in the rear cockpit. The Department of Defense estimated that the use of navigators rather than pilots would result in savings exceeding $400 million over a 5-year period.

The Air Force was concerned that such a change in training and assignment policies would have a detrimental impact on combat effectiveness and aircraft safety. The rationale for the use of pilots in the rear cockpit of the F-4 is based on their ability to serve as backup pilots to provide occasional relief for the aircraft commander, to recover the aircraft from dangerous attitudes if the aircraft commander experiences severe spatial disorientation, and to fly the aircraft if the aircraft commander becomes incapacitated. Both pilots and navigators receive navigational training while only pilots receive formal training in handling aircraft controls and flight dynamics. Therefore, a pilot has been considered the best choice to operate the rear cockpit systems and provide support to the aircraft commander.

Most aircraft controls are duplicated in both cockpits, but controls of certain navigational systems, such as radar and the inertial navigation system, are located only in the rear cockpit. The basic requirement for a second-seater is to operate the navigational subsystems effectively and to be sufficiently familiar with aircraft dynamics and mission requirements to provide timely support to the aircraft commander.

Although most rear-cockpit pilots eventually upgrade to the position of aircraft commander, the period of "apprenticeship" for a second-seater has averaged two to three years. Some pilots have expressed the opinion that serving for a period of time in an essentially nonpilot capacity may be detrimental to pilot skills, motivation, and overall job performance.

Operational Problem

In response to the DOD request, Headquarters, United States Air Force directed the Armament Development and Test Center, Eglin Air Force Base, to conduct a test in both the United States and Southeast Asia to determine if there are significant differences in the combat performance of navigator and pilot second-seaters that would preclude the assignment of navigators to the F-4 or other two-place fighter aircraft. The test was named Project Combat Team.

Technical Problem

Overall performance in the highly complex, high-speed environment of F-4 combat is multidimensional in nature. Thus, analysis of data from several behavioral parameters is required. These parameters include differences in the frequency with which tasks are performed and the amount of time spent on the various tasks, proficiency in the accomplishment of certain basic tasks and duties, and the individual second-seater's ability to provide timely and appropriate support to the aircraft commander.

One of the objectives of Project Combat Team was to measure the combat proficiency of newly graduated pilots and navigators who were assigned to the second-seat positions in F-4 aircraft. Attention was focussed on several specific questions: whether there were significant differences between the two groups in the proficiency with which they performed certain inflight tasks and duties during combat missions; if there were differences, whether they were maintained as combat experience increased; and whether there were measurable differences between the two groups in accomplishing mission success or maintaining aircraft safety.

A contaminating variable in assessing proficiency of second-seaters was recognized in the fact that the performance of the aircraft commander is the primary factor in accomplishing mission objectives, responding to environmental threats, and maintaining aircraft safety. The basic requirement for the second-seater is to operate the rear-cockpit subsystems effectively and to be sufficiently familiar with aircraft dynamics and mission requirements to provide timely support to the aircraft commander.

Since the design of the study called for the use of assessments and ratings, another methodological concern was related to the psychometric characteristics of ratings and rating scales. Proficiency ratings of various types have provided much of the information required for the operation of many personnel systems. The utility of proficiency

ratings, evidenced by their widespread use for such personnel actions as promotion, career progression, and assignments, is in part offset by well documented weaknesses and defects (Tiffin & McCormick, 1965). The recognized deficiencies of rating scales include a tendency toward inflation of ratings with repeated use, contamination by halo effects, and lack of uniformity in standards between different raters. Such pervasive defects have made it difficult to make inter-group comparisons.

Ratings of performance, however, have certain compensating advantages (Guilford, 1954). Studies have shown that most performance ratings have a satisfactory degree of validity. Furthermore, ratings may be used to assess job proficiency in situations where adequate objective criteria of performance are lacking. Ratings are also relatively easy to administer. Sources of error, such as the halo effect, can be reduced through careful planning of the rating scale format and administration procedures. Finally, it was anticipated that aircraft commanders would be satisfactory raters because of their qualifications as experienced F-4 crewmembers and their ability to monitor and evaluate the second-seater's proficiency.

Scope

A number of assessment instruments were developed as a part of Project Combat Team to provide comprehensive data on aircrew behavior and performance during combat missions. These included an inflight task inventory, a proficiency rating scale, a mission background summary, and a standardized post-mission interview. This report presents the data obtained from the proficiency rating scale. Mission evaluation data drawn from the operational reports filed after each combat mission are also presented. In addition, the role of proficiency ratings in contributing to an overall assessment of operationally significant differences between pilot and navigator rear-seat crewmembers is discussed. A discussion of the F-4 task inventory is presented in an earlier report (Ratliff, Shore, Chiorini, & Curran, 1969).

II. METHOD

Subjects

After completion of their respective undergraduate pilot and navigator training courses, the test subjects, 12 pilots and 12 navigators, were assigned to F-4 combat crew training at Davis-Monthan Air Force Base. Both the pilots and navigators were selected from recent undergraduate pilot and navigator course graduates available for assignment.

Normally, 60 training missions are given to each crew at Davis-Monthan. A crew consists of a newly upgraded aircraft commander and a new second-seater assigned to fly with him in Southeast Asia. Crewmembers are normally trained together as a team to provide crew integrity. Since only second-seaters were subjects in the present project, the combat crew training course was reduced to 30 missions during which second-seaters were trained by an experienced instructor pilot in the front seat on all flights. The added experience of the instructor pilots was assumed to compensate for the elimination of orientation flights and missions which were primarily used as training flights for the student aircraft commander.

Data Collection Forms

The data collection instruments were developed at Davis-Monthan Air Force Base based on information obtained during a series of interviews with 15 F-4 instructor pilots, most of whom had Southeast Asia combat experience. These interviews were conducted to determine those inflight crew functions in which second-seat proficiency affects crew performance or mission effectiveness, and to identify significant inflight F-4 crew tasks. The data collection forms provided for a proficiency evaluation of the second-seater, a mission evaluation, and an operational report of the mission.

AFPT 22-2, F-4 Second-Seater Proficiency Evaluation. This instrument, shown in Appendix I, was designed to measure proficiency in ten second-seat functions and three general performance characteristics. The rating categories included operation of aircraft systems, crew coordination, and personal characteristics. Aircraft commanders were instructed to compare the performance of test second-seaters with that of other second-seaters with similar experience. These forms were administered on a trial basis for a series of ten training missions to the second-seaters and their instructor pilots at Davis-Monthan. Following this experimental use of the forms, the rating categories were revised on the basis of an assessment of crew responses and post-flight crew interviews. In addition, the rating scale was expanded from 7 to 11 points, a change which improved the reliability of the instrument.

AFPT 22-1, Section 3, Mission Evaluation. The mission evaluation form is shown in Appendix I.

Its purpose was to obtain an evaluation from both the aircraft commander and the second-seater of the degree to which mission objectives were achieved and to record the frequency and circumstances in which the second-seater made outstanding contributions to or detracted from mission effectiveness. Post-mission interviews were conducted to clarify or expand any information recorded on the form.

OP REP 4, Operational Report. This document contained mission data that could be used to evaluate the extent of damage to enemy defenses, supply lines, etc. The report was completed after each mission during the standard intelligence debriefing given to all returning crews.

Data Collection Procedures

Two Southeast Asia tactical fighter wings, the 8th TFW, Ubon RTAB, and the 366th TFW, Da Nang RVNAB, were each assigned a group of 12 test second-seaters, 6 pilots and 6 navigators. Two personnel measurement psychologists were assigned to each base to collect data using standardized collection procedures. Prior to data collection, the aircraft commanders were extensively briefed on the criteria which were to be used in evaluating the performance of the second-seater. The Mission Evaluation form (AFPT 22-1, Section 3) was administered to both crewmembers immediately following each combat mission. Both the second-seater and the aircraft commander completed these forms within an hour of landing. The aircraft commanders were then interviewed to amplify information concerning the mission and the second-seater's performance. Each mission was treated as a separate data entry. Information contained in the Operational Report (OP REP 4) supplemented the data in the completed Mission Evaluation. After each block of five missions flown by a test second-seater, the F-4 Second-Seater Proficiency Evaluation (AFPT 22-2) was administered to each aircraft commander who flew with him during that block of missions.

In completing a proficiency rating, the aircraft commanders were instructed to compare the second-seater being rated with second-seaters of equal experience in terms of the number of combat missions flown. No entries were made for categories in which performance was not observed.

Both the aircraft commanders and the second-seaters were assured that the information obtained on the various data collection forms would be handled in strictest confidence and would not become a part of any personnel records. Complete data were collected for the first 30 to 35 combat missions on 22 of the 24 test second-seaters. One pilot was listed as missing in action after his 17th mission and one navigator reached Southeast Asia too late to complete 30 missions by the data cutoff date.

Differences between pilots and navigators on each rating category were tested by a multiple linear regression analysis (Bottenberg & Ward, 1963). Three specific questions were tested concerning each category: (*a*) Were there proficiency differences between pilot and navigator second-seaters during the first 30 combat missions? (*b*) Were the magnitudes of any differences stable as combat experience increased? (*c*) Were there any proficiency differences between pilots and navigators by the time they had flown from 26 to 30 missions?

Mission Evaluation

A board of 12 aircraft commanders who were currently assigned to combat units in Southeast Asia was convened at Eglin Air Force Base to rate the effectiveness of combat missions in which the test second-seaters participated in Southeast Asia. The board was composed of ten majors and two lieutenant colonels with recent combat experience in the F-4 ranging from 50 to 150 missions. Information available to the board for each mission rated included second-seater and aircraft commander responses to the mission evaluation portion of AFPT 22-1, a copy of the Operational Report, and copies of post-mission interviews with the aircraft commanders. The board had no information as to crew composition for any of the missions rated.

The 12 board members were divided into two working groups and instructed to rate the missions according to the extent to which they were accomplished in accordance with the target description appearing on the Operational Report. They worked individually and rated the missions in terms of four possible alternatives: all mission objectives achieved, some mission objectives achieved, no mission objectives achieved, or insufficient information to rate the mission. If fewer than four of the six judges agreed upon a rating, the mission was discussed by the six-man group. If agreement was not reached in this conference, the mission was rated as non-scorable.

Data Restrictions

Restriction of missions to South Vietnam and the southern portion of North Vietnam during the

entire Southeast Asia data collection period limited the spectrum of missions upon which data could be collected. Missions over the most heavily defended areas of North Vietnam and observations of the test subjects reacting to MIG or SAM threats were not available for this report. Any interpretation of the data should take these limitations into account.

III. RESULTS

The results of the second-seater proficiency ratings made by the aircraft commanders are shown in Figure 1. The results of the statistical analyses are presented in Appendix II. Navigators were rated substantially higher than pilots on Category 2, Use of Radar ($p < .01$). This difference was considerably reduced by the 26-to-30 mission segment, and group differences at this point were not statistically significant. Pilots were rated as more proficient than navigators on Category 5, Visual Target Acquisition and Identification ($p < .03$). Differences at the 26-to-30 mission segment were not statistically significant. Pilots were also rated higher ($p < .07$) than navigators on Category 4, Understanding Radio Communications, and this group difference was significant at the 26-to-30 mission segment ($p < .04$). Category 7, Calling Out Correct Action in Response to a Threat, and Category 8, Emergency Situations, were not analyzed since second-seaters were infrequently rated on these categories.

An estimate of the reliability of the ratings was calculated by correlating the mean of the first, third, and fifth ratings on each category with the mean of the second, fourth, and sixth ratings. These figures are presented in Table 1. Categories 7 and 8 are not presented due to insufficient data.

In order to determine the sensitivity of the measurement instrument, separate computations of means were made for pilots and navigators by base of assignment. Whereas pilots had been assigned to the two Southeast Asia bases without regard to standings in their combat crew training class, navigators with the highest class standings were assigned to Ubon, and those with the lowest were assigned to Da Nang. The method of assignment provided an opportunity to observe the test instrument's sensitivity in detecting the difference in the two navigator groups. Figures 2 and 3 present the mean proficiency ratings by base. Categories 7 and 8 are not presented due to insufficient data. Ubon navigators were rated superior to Da Nang navigators on all categories, and the differences remained consistent for the entire data collection period. On the other hand, there were no consistent differences between the two groups of pilots.

Additional information concerning second-seater proficiency was obtained from the Mission Evaluation, AFPT 22-1, Section 3. In question 4, both the second-seater and his aircraft commander were asked to identify second-seater actions which required corrective action by the aircraft commander. These actions included second-seater errors in equipment operation, failure to make radio frequency changes, etc. In question 5, information was requested from both the second-seater and his aircraft commander concerning instances during the mission when the second-seater provided outstanding support to the aircraft commander. Comments here ranged from unqualified statements (e.g., "excellent support") to specific instances, such as outstanding use of particular pieces of equipment under stressful conditions. Responses to question 4 were tallied when either the aircraft commander or the second-seater responded, whereas only the aircraft commanders' responses to question 5 were scored. The data from questions 4 and 5 are presented in Table 2.

Both pilots and navigators showed decreases over time in the percentage of missions on which their actions required correction or compensation by their aircraft commanders. Pilots showed a more rapid decrease and had an overall lower percentage of such missions. Responses to question 5 showed little differentiation between the two groups, and there were no consistent changes in the percentages of favorable comments over time.

The comparisons between pilots and navigators on mission effectiveness over all 30 missions made by the board of 12 aircraft commanders are summarized in Table 3. Approximately the same number of missions was scored for both pilots and navigators. The data were analyzed in two categories: all mission objectives achieved, and less than all objectives achieved. Pilots showed significantly more successful and fewer unsuccessful missions than did navigators ($p < .05$).

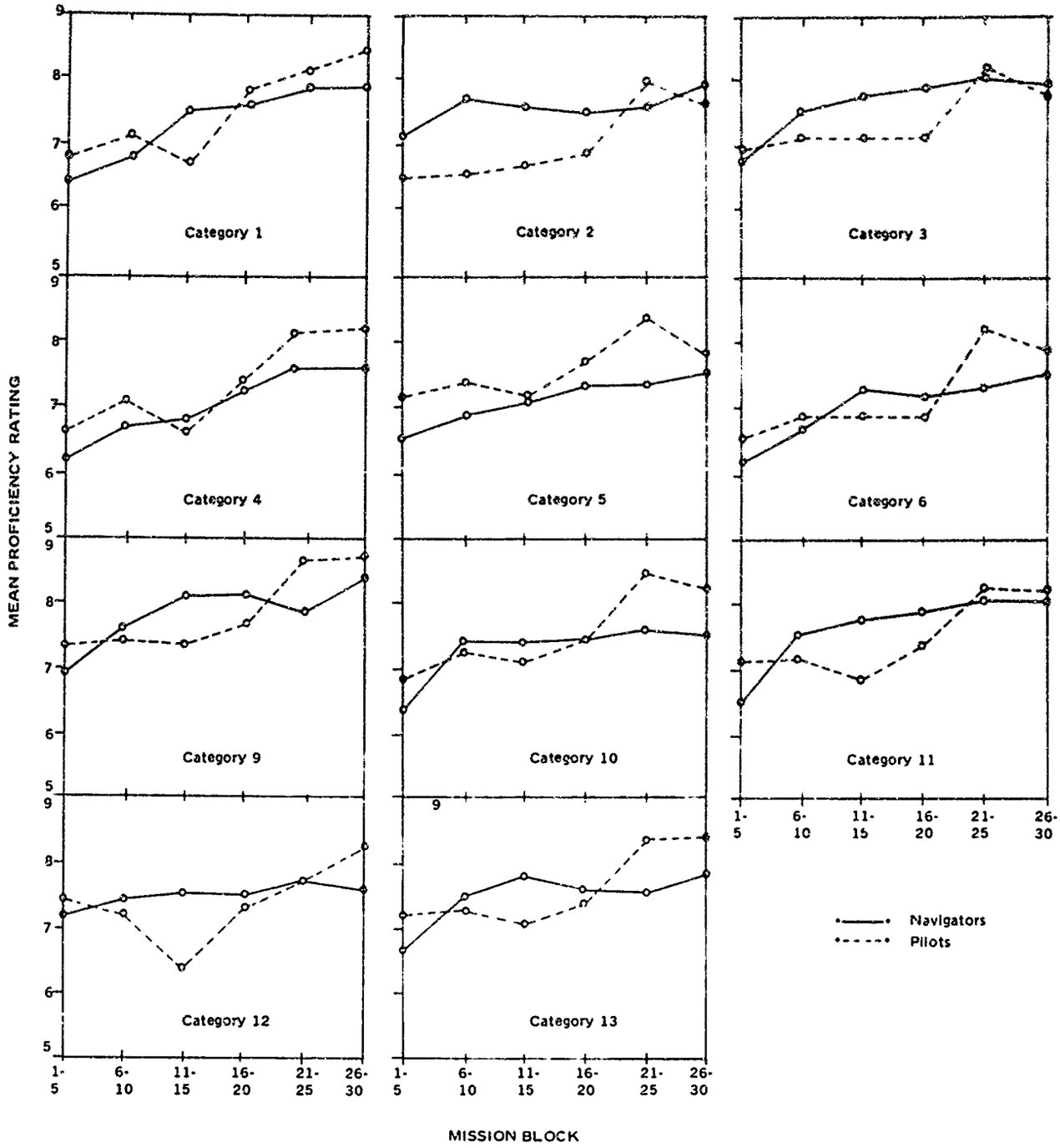


Fig. 1. Pilot and navigator second-seater group ratings on proficiency categories, averaged over 5-mission blocks.

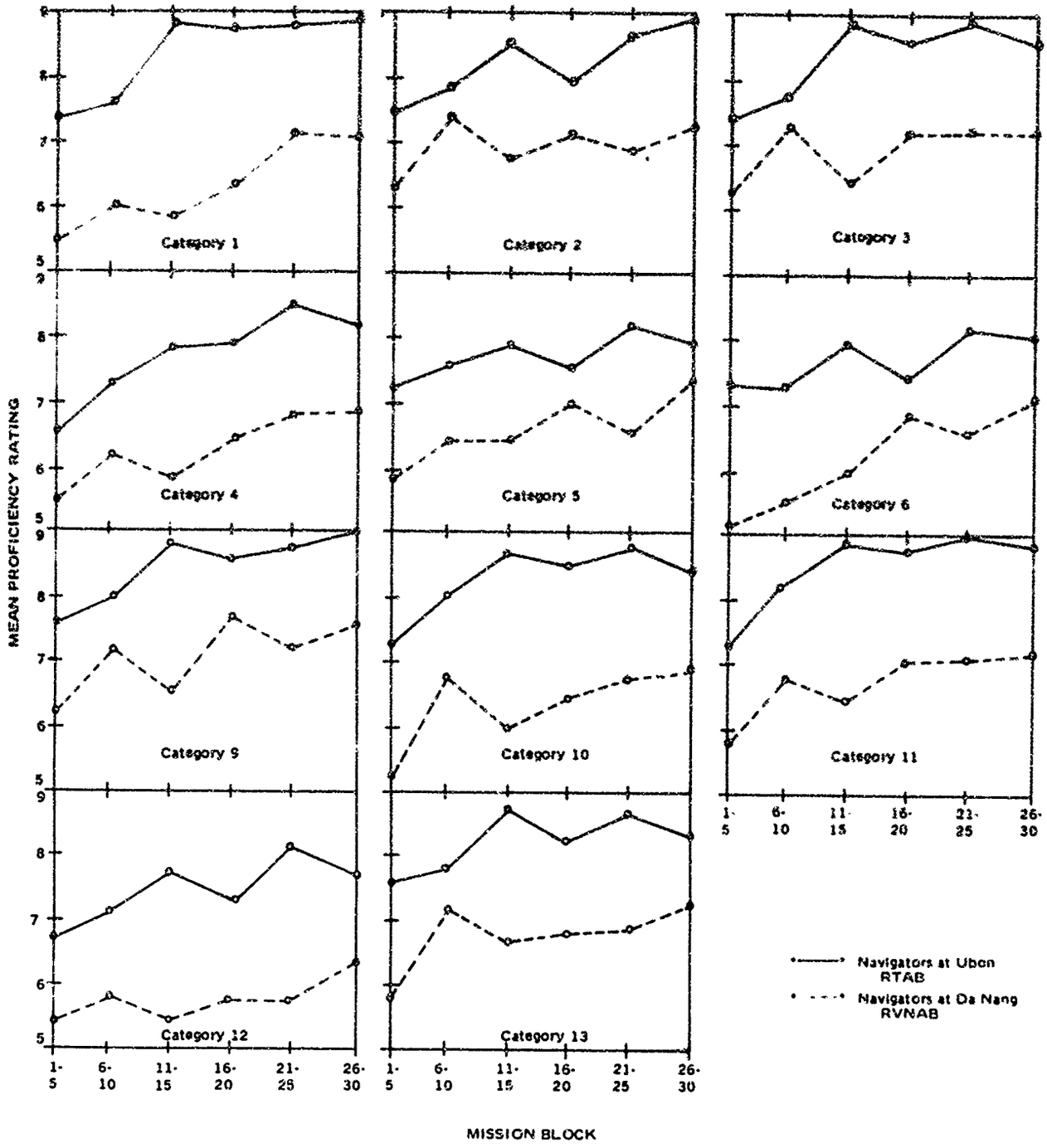


Fig. 2. Navigator second-seater group ratings on proficiency categories by base of assignment, averaged over 5-mission blocks.

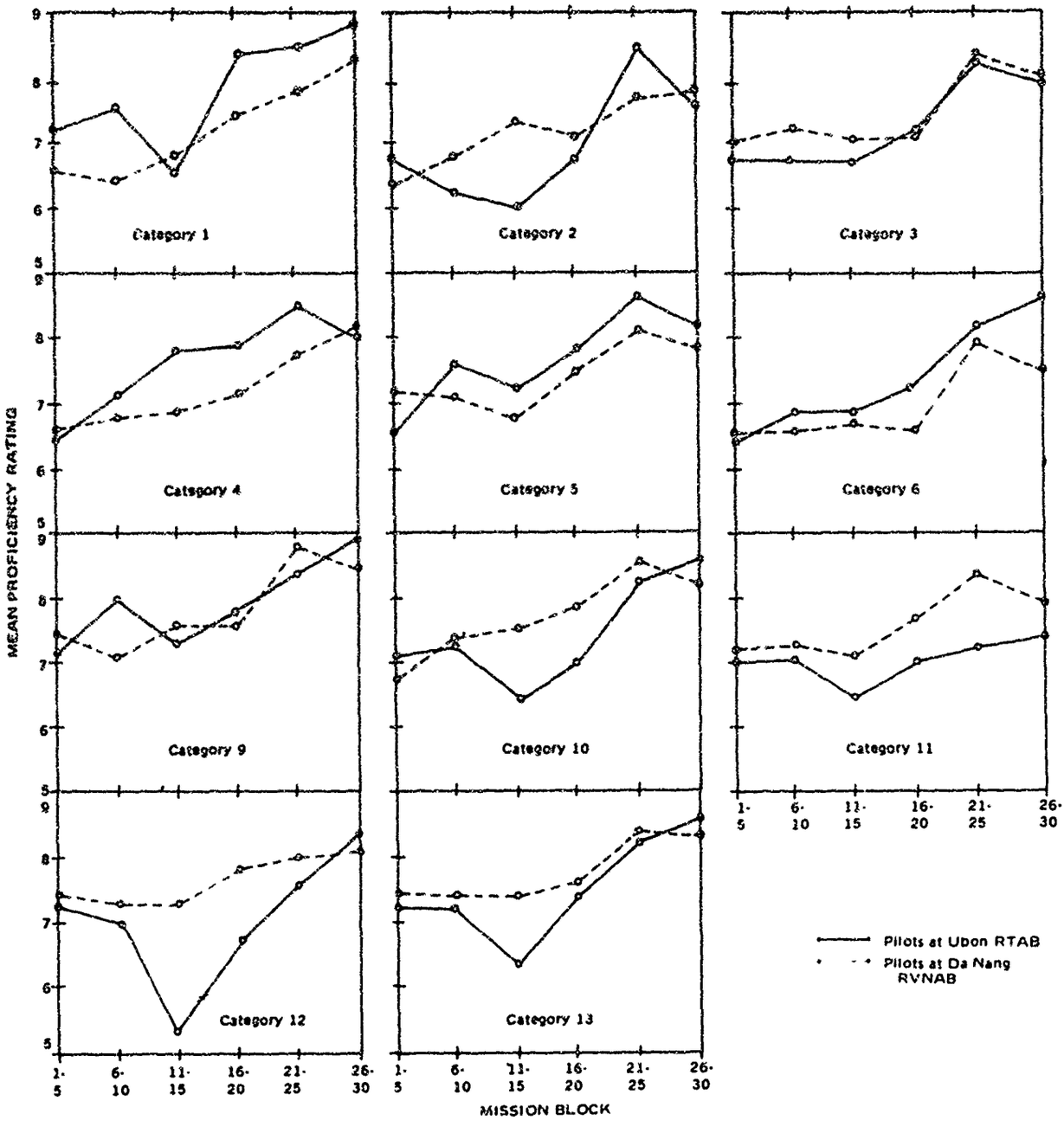


Fig. 3. Pilot second-seater group ratings on proficiency categories by base of assignment, averaged over 5-mission blocks.

Table 1. Reliability Coefficients for Rating Categories of Second-Seater Proficiency Evaluation

(Rating Categories from AFFT 22-2, F-4 Second-Seater Proficiency Evaluation)

	Rating Category	Reliability Coefficient
1	Mission Preparation	.83
2	Use of Radar	.78
3	Navigation	.81
4	Understanding Radio Communications	.70
5	Visual Target Acquisition and Identification	.48
6	Visual Defensive Surveillance	.67
7[a]	Calling Out Correct Action in Response to a Threat	-
8[a]	Emergency Situations	-
9	Reaction to Combat Stress	.47
10	Staying Ahead of Situation and Keeping Aircraft Commander Informed During Mission	.67
11	Job Knowledge	.69
12	Attitude Toward Job Performance	.47
13	Overall Performance	.34

[a]Reliabilities for Categories 7 and 8 were not computed because of insufficient data.

Table 2. Percentage of Missions on Which the Second-Seater's Actions Detracted from or Contributed Outstandingly to Mission Effectiveness

(Computed from Responses to Questions 4[a] and 5[b] on the Mission Evaluation, AFPT 22-1, Section 3.)

Mission When Action Occurred	Percentage of Missions When Action Occurred	
	Navigators	Pilots
Actions Required Compensation		
Missions 1 -10	20.8	17.5
Missions 11-20	17.5	6.8
Missions 21-30	13.3	4.5
All Missions Combined	17.2	9.8
Actions Provided Outstanding Support		
Missions 1 -10	30.0	25.0
Missions 11-20	25.0	20.0
Missions 21-30	30.0	39.1
All Missions Combined	28.3	28.0

[a]Question 4. Did second-seat crewmember's performance detract from mission effectiveness or require compensation by the aircraft commander?

[b]Question 5. Were there instances when the second-seat crewmember provided outstanding support to the aircraft commander or to mission effectiveness?

Table 3. Mission Effectiveness Judgments by Board of Aircraft Commanders

(Judgments Based on Data from the Mission Evaluation, AFPT 22-1, Section 3)

Mission Effectiveness Evaluation	Missions with Navigator Second-Seaters		Missions with Pilot Second-Seaters	
	N	(Percent)	N	(Percent)
All Objectives Achieved	305	(87)	319	(92)
All Objectives Not Achieved	45	(13)	27	(8)

$\chi^2 = 4.79$

$p < .05$

IV. DISCUSSION

The proficiency ratings showed no general superiority in performance of one group of second-seaters over the other. Specifically, navigators received higher proficiency ratings than pilots in the Use of Radar while pilots received higher ratings in Understanding Radio Communications and Visual Target Acquisition. However, the important finding for operational consideration is that, of the few significant differences found between pilot and navigator second-seaters, only in Understanding Radio Communications was a difference maintained to the 26-to-30 mission segment.

If the observed proficiency differences between pilot and navigator second-seaters had been maintained over the 30 missions sampled, such differences might have been related either to training differences or to selection differences. However, since the proficiency differences decreased sharply over the first 30 combat missions, it can be concluded that proficiency differences between pilots and navigators can be reduced or eliminated as a result of modifications in crew training programs.

The method used to inspect the sensitivity of the ratings compares the ratings obtained in combat with rankings made earlier for performance during training. Specifically, the earlier rankings were the class standings of the second-seaters at the completion of combat crew training. Navigators assigned to Ubon were all ranked higher than navigators assigned to Da Nang, and their combat proficiency ratings consistently discriminated between these two groups. Pilots at the two bases were similar in terms of class standings and combat proficiency ratings. Thus, when real differences were known to exist in the proficiency of second-seaters, differences appeared in the proficiency ratings which were consistent with predictions that could be made from the nature of the previously known group disparities. To the extent that the class standings from combat crew training are a valid measure of performance, then the group performance curves by base provide an index of the validity of the rating categories used in the proficiency evaluation instrument (AFPT 22-2).

The reliability of the proficiency data was estimated by calculating reliability coefficients comparing the mean of the odd- and even-numbered mission blocks. Because the proficiency of both the pilots and the navigators was increasing over the first 30 missions, the reliability coefficients calculated for each rating category were considered to be conservative. However, the transitional nature of the behavior being rated in the present study made more traditional methods for computing reliability coefficients less appropriate. Other factors which may have influenced the reliability coefficients include the large number of judges that had to be used to rate the test second-seaters, the wide range of mission requirements between sorties, and the large variations in combat stress from mission to mission.

The validity and reliability demonstrated in the data answered a basic question raised by Project Combat Team. Namely, can objective, reliable proficiency ratings be collected when complex interpersonal relationships are established between rater and ratee in a highly stressful environment? Since the observed psychometric characteristics of the ratings obtained in this study were comparable to those of rating data collected under laboratory conditions (Guilford, 1954, Tiffin & McCormick, 1965), it appears that it is possible to collect useful ratings of proficiency under combat conditions.

Previous work with rating scales has concentrated on the analysis of task or job proficiency for individuals or for homogeneous groups of individuals. The present data represent a significant departure from previous work in that the job was held generally constant while the proficiency of two groups with known differences in their training backgrounds was inspected. Thus, it was possible to inspect the differences in proficiency at a certain job (F-4 second seat) between two groups in a strictly operational setting. The unique experimental plan for Project Combat Team, requiring multidimensional behavioral data, also provided an opportunity to compare the results of the proficiency ratings of the two groups with other performance data obtained on these groups using a task inventory. Inconsistencies in the results of these two performance evaluations would have weakened confidence in the data from either or both instruments. However, the findings of the proficiency ratings of pilot and navigator second-seaters were completely consistent with the findings of the task inventory data. A more detailed discussion may be found in an earlier report on the study (Ratliff *et al.*, 1969).

The data collected from post-mission interviews (AFPT 22-1, Part 3) and intelligence debriefing reports (OI REP 4) showed that, over the first 30 combat missions, pilot second-seaters flew on more missions rated completely successful and fewer rated as not meeting all the mission objectives than did the navigators; the differences were small but statistically significant.

The specific contribution to mission effectiveness of those tasks for which proficiency ratings were made could not be objectively evaluated. Thus, the judgment of individuals responsible for operational decision-making at the upper levels of management must be relied upon to identify the causal linkage between proficiency at certain inflight tasks and combat mission success. A discussion of the possible ways the proficiency, mission evaluation, and task analysis data collected for Project Combat Team may be integrated by management personnel can be found in Ratliff *et al.* (1969).

The reliability of the proficiency data may have been improved if the proficiency evaluation form had been administered during the post-mission interview immediately after each mission rather than after each block of five missions. The latter procedure was established, however, to minimize the data collection time required of crewmembers immediately following each combat mission. Another way in which the psychometric characteristics of proficiency data collected under combat conditions may be improved is to include the proficiency categories at appropriate locations in the task analysis form. This can be done without appreciably lengthening the amount of time taken by the post-mission interview. In a study which tested such a procedure (Tiffin & McCormick, 1965), it was found that proficiency information collected in this manner had higher coefficients of reliability than proficiency data collected under identical conditions on a separate form.

V. SUMMARY AND CONCLUSIONS

In response to a request from the Department of Defense, the Air Force established Project Combat Team to study the feasibility of replacing pilots with navigators in the rear cockpit of F-4 fighter aircraft. In the absence of objective performance criteria in the combat setting, it was determined that data should be collected using several performance parameters. For this study, data were collected to answer three questions: (*a*) Are there significant differences between pilot and navigator second-seaters in the proficiency with which they perform certain inflight tasks and duties during combat missions? (*b*) If found, are differences in proficiency maintained as combat experience increases? (*c*) Do differences in training background result in measurable differences in mission success or aircraft safety? Since aircrew proficiency had never been measured under combat conditions, the data were also analyzed to determine if it is feasible to collect reliable proficiency measures in a combat environment.

Analyses of the proficiency rating information included pilot and navigator group means for six successive five-mission blocks on 13 proficiency categories, group mean data plotted as a function of Southeast Asian base of assignment, and estimates of test-retest reliability. Mission evaluation data and operational report information generated during the standard intelligence debriefing were also presented in order to inspect the success of combat missions flown with pilots or navigators occupying the second seat of the aircraft.

These data, although collected in a limited combat environment, lead to several conclusions with regard to some methodological questions involving the data collection instruments, as well as to the more specific questions of pilot and navigator performance differences.

1. It is possible to obtain proficiency data in a combat environment that are sufficiently reliable to be used by higher management in determining the crew manning requirements of F-4 aircraft.

2. It is possible to use ratings of task or job proficiency to measure the differences between two groups in an operational setting.

3. The proficiency rating scale used in the present study was sensitive enough to reflect group disparities which were previously known to exist in the pilot and navigator subjects.

4. The proficiency ratings showed no general superiority in performance of one group of second-seaters over the other.

5. Early in their combat experience, navigators received higher proficiency ratings in the Use of Radar while pilots received higher ratings in Understanding Radio Communications and Visual Target Acquisition. However, the important operational finding was that these early differences were sharply reduced by the time they had flown 26 to 30 missions, except in the case of Understanding Radio Communications.

6. The transitional nature of the proficiency differences indicates that they are a function of pilot and navigator training differences and can, therefore, be reduced or eliminated through modifications in crew training programs.

7. During the first 30 combat missions, pilot second-seaters flew on more missions rated completely successful than did navigators; the differences were small but statistically significant. However, these differences in mission success could not be objectively related to proficiency differences between pilots and navigators.

REFERENCES

Bottenberg, R.A. & Ward, J.H., Jr. *Applied multiple linear regression* PRL-TDR-63-6, AD413 128. Lackland AFB, Tex.: Personnel Research Laboratory, Aerospace Medical Division, March 1963.

Guilford, J.P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

Ratliff, F.R., Shore, C.W., Chiorini, J.R., & Curran, C.R. *Inflight performance differences of pilot and navigator F-4 second-seat crewmembers: A limited Southeast Asia combat evaluation.* AFHRL-TR-69-104. Lackland AFB, Tex.: Personnel Research Division, Air Force Human Resources Laboratory, July 1969.

Tiffin, J. & McCormick, E.J. *Industrial psychology.* (5th ed.) Englewood Cliffs, New Jersey: Prentice-Hall, 1965.

BLANK PAGE

APPENDIX I. DATA COLLECTION INSTRUMENTS[1]

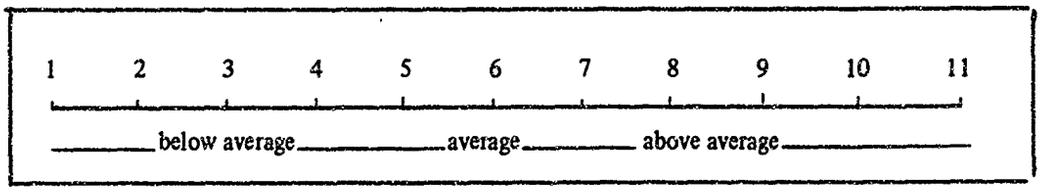
DO NOT WRITE IN THESE SPACES

GIB	A/C	WHO	#GIB MSNS	CELL

F-4 GIB
PROFICIENCY EVALUATION

A/C: Consider the GIB's performance on all missions he has flown with you since your last evaluation of him. Rate him relative to other GIBs with a similar number of combat missions.

GIB: Consider your own performance on all missions since you last made this evaluation. Rate yourself relative to other GIBs with a similar number of combat missions.



Enter the appropriate number for a GIB rating on each of the following factors.

RATE PERFORMANCE IN: (Consider factors observed since last evaluation)

1. Mission preparation . . . . . _____
2. Use of radar . . . . . _____
3. Navigation . . . . . _____
4. Understanding radio communications . . . . . _____
5. Visual target acquisition and identification . . . . . _____
6. Visual defensive surveillance . . . . . _____
7. Calling out correct action in response to a threat . . . . . _____
8. Emergency situations . . . . . _____
9. Reaction to combat stress . . . . . _____
10. Staying ahead of the situation and keeping the A/C informed during mission . . . . . _____

[1]The term GIB, or "Guy in back," used in the data collection instruments is the operational vernacular for the second-seat crewmember.

F-4 GIB Proficiency Evaluation *(Continued)*

RATE GIB ON THESE GENERAL FACTORS:

1. JOB KNOWLEDGE . . . . . _____

2. ATTITUDE TOWARD JOB PERFORMANCE . . . . . _____

* * * * * * * * * * * * * * * * * * *

RATE GIB's OVERALL PERFORMANCE . . . . . _____

COMMENTS: (Strengths, weaknesses, other)

DO NOT WRITE IN THESE SPACES

GIB	A/C	WHO	MSN #	PKG #	MSN TYPE

MISSION EVALUATION

The answers to the following questions will be kept strictly confidential. It is important that the questions be answered with complete frankness. This material will NOT be reflected in any report about a particular GIB or A/C but will be essential in helping us achieve the object of our study.

1. If the GIB performed flying duties during any part of the mission (flight formation, recovery from unusual attitude, combat maneuvers, defensive maneuvers in response to MIGs, SAMs, and AAA, letdown, etc.), please describe the situation and tell why GIB handled the flight controls.

_____

_____

2. Rate effectiveness of mission as planned:

( ) a. All objectives achieved.
( ) b. Some objectives achieved.
( ) c. No objectives achieved.
Explain b or c:

_____

_____

3. Rate effectiveness of mission as it was flown: (Consider changes in plans as mission progressed).

( ) a. All objectives achieved.
( ) b. Some objectives achieved.
( ) c. No objectives achieved.
Explain a, b, or c: _____

_____

4. Did GIB's performance detract from mission effectiveness or require compensation by the A/C?

( ) No
( ) Yes - Briefly explain: _____

_____

5. Were there instances when the GIB provided outstanding support to the aircraft commander or to mission effectiveness?

( ) No
( ) Yes - Briefly explain: _____

_____

6. COMMENTS: _____

_____

_____

APPENDIX II. MULTIPLE LINEAR REGRESSION ANALYSIS OF PROFICIENCY DATA

PREDICTOR VARIABLES

For the regression analyses, 34 predictor variables were generated. Equations were solved for each of the performance variables using various combinations of these predictors.

Predictor Variable	Description	Score Range
X_1	Pilot	
X_2	Navigator	0,1
X_3	1-5 Mission Block	0,1
X_4	6-10 Mission Block	0,1
X_5	11-15 Mission Block	0,1
X_6	16-20 Mission Block	0,1
X_7	21-25 Mission Block	0,1
X_8	26-30 Mission Block	0,1
X_9	Pilot x 1-5 Mission Block	0,1
X_{10}	Pilot x 6-10 Mission Block	0,1
X_{11}	Pilot x 11-15 Mission Block	0,1
X_{12}	Pilot x 16-20 Mission Block	0,1
X_{13}	Pilot x 21-25 Mission Block	0,1
X_{14}	Pilot x 26-30 Mission Block	0,1
X_{15}	Navigator x 1-5 Mission Block	0,1
X_{16}	Navigator x 6-10 Mission Block	0,1
X_{17}	Navigator x 11-15 Mission Block	0,1
X_{18}	Navigator x 16-20 Mission Block	0,1
X_{19}	Navigator x 21-25 Mission Block	0,1
X_{20}	Navigator x 26-30 Mission Block	0,1
X_{21}	Pilot x 1-5 Mission Block + 2(Pilot x 6-10 Mission Block) + 3(Pilot x 11-15 Mission Block) + 4(Pilot x 16-20 Mission Block) + 5(Pilot x 21-25 Mission Block) + 6(Pilot x 26-30 Mission Block)	
X_{22}	Navigator x 1-5 Mission Block + 2(Navigator x 6-10 Mission Block) + 3(Navigator x 11-15 Mission Block) + 4(Navigator x 16-20 Mission Block) + 5(Navigator x 21-25 Mission Block) + 6(Navigator x 26-30 Mission Block)	0,1,. . .,21
X_{23}	$(X_{21})^2$	0,1,. . .,441
X_{24}	$(X_{22})^2$	0,1,. . .,441
X_{25}	$X_{21} + X_{22}$	0,1,. . .,42
X_{26}	$X_{23} + X_{24}$	0,1,. . .,882
X_{27}	$X_1 + X_{22}$	0,1, . . .,22
X_{28}	$X_1 + X_{24}$	0,1,. . .,442
X_{29}	$X_{21} - X_1$	0,1,. . .,20
X_{30}	$X_{23} - X_1$	0,1,. . .,440
X_{31}	$X_{22} + 6X_1$	0,1,. . .,27
X_{32}	$X_{24} + 36X_1$	0,1,. . .,447
X_{33}	$X_{21} - 6X_1$	0,1,. . .,15
X_{34}	$X_{23} - 36X_1$	0,1,. . .405

HYPOTHESES

Of the various hypotheses tested, six were of particular interest in this study. For each of the performance variables, the following questions were analyzed to assess pilot and navigator performance differences.

1. Are the magnitudes of differences between pilot and navigator second-seat crewmembers stable as combat experience increases? (Analysis using common weight on linear and quadratic cell number vectors)

2. During the first 30 missions, is there an overall difference between the proficiency of pilot and navigator second-seat crewmembers? (Analysis using both linear and quadratic weights on cell number vectors)

3. Are the magnitudes of differences between pilot and navigator second-seat crewmembers stable as combat experience increases? (Analysis using common weight on linear cell number vectors)

4. During the first 30 missions, is there an overall difference between the proficiency of pilot and navigator second-seat crewmembers? (Analysis using linear weights on cell number vectors)

5. Are there proficiency differences between pilot and navigator second-seat crewmembers during their first five combat missions?

6. Are there proficiency differences between pilot and navigator second-seat crewmembers during their 26th to 30th missions?

RESULTS

Results of the analyses for the performance variables are presented in Table 4. For each variable (i.e., proficiency category), results of the F tests for the six hypotheses are presented.

Table 4. Results of Multiple Regression Analysis and F Tests of Significance for Proficiency Data

Hypothesis	R^2 Full Model	R^2 Restricted Model	df_1	df_2	F	Significance Level
Category 1: Mission Preparation						
1	.1633	.1493	2	131	1.0911	.33889
2	.1493	.1457	1	133	0.5685	.45220
3	.1502	.1493	1	133	0.1388	.71003
4	.1493	.1457	1	134	0.5726	.45055
5	.1633	.1572	1	131	0.9500	.33151
6	.1633	.1495	1	131	2.1633	.14373
Category 2: Use of Radar						
1	.1564	.1364	2	130	1.5382	.21865
2	.1364	.0921	1	132	6.7664	.01035
3	.1515	.1362	1	132	2.3761	.12560
4	.1362	.0920	1	133	6.8167	.01007
5	.1564	.1373	1	130	2.9442	.08858
6	.1564	.1554	1	130	0.1496	.69953
Category 3: Navigation						
1	.1302	.1141	2	131	1.2132	.30055
2	.1141	.1000	1	133	2.1218	.14758
3	.1122	.1115	1	133	0.1084	74246
4	.1115	.0972	1	134	2.1582	.14415
5	.1302	1298	1	131	0.0637	.80114
6	.1302	.1285	1	131	0.2562	61358
Category 4: Understanding Radio Communications						
1	.2070	.1931	2	131	1.1450	.32138
2	.1931	.1732	1	133	3.2819	.07231
3	.1948	1928	1	133	0.3352	.56357
4	1928	.1728	1	134	3.3202	.07065
5	.2070	1967	1	131	1.6958	.19511
6	2070	1816	1	131	4.1904	.04265
Category 5: Visual Target Acquisition and Identification						
1	1096	1084	2	129	0 0859	91772
2	1084	0765	1	131	4 6936	03208
3	1083	1071	1	131	0 1790	67290
4	1071	0752	1	132	4 7150	03169
5	1096	1050	1	129	0 6600	41808
6	1096	0976	1	129	1 7439	18898
Category 6: Visual Defensive Surveillance						
1	1536	1447	2	118	0 6205	53945
2	1447	1401	1	120	0.6457	42323
3	1457	1441	1	120	0.2285	.63352
4	1441	1394	1	121	0.6597	41828
5	1536	1506	1	118	0.4168	51977
6	1356	1418	1	118	1 6465	20194

Table 4 *(Continued)*

Hypothesis	R^2 Full Model	R^2 Restricted Model	df_1	df_2	F	Significance Level
Category 7: Calling Out Correct Action in Response to a Threat						
1	.2238	.0672	2	96	3.5043	.03397
2	.1672	.1434	1	98	2.7919	.09794
3	.1703	.1635	1	98	0.7988	.37365
4	.1635	.1388	1	99	2.9240	.09041
5	.2238	.2014	1	96	2.7686	.09939
6	.2238	.1569	1	96	8.2826	.00493
Category 8: Emergency Situations						
1	.1174	.1160	2	77	0.0593	.94242
2	.1160	.0944	1	79	1.9346	.16816
3	.0713	.0706	1	79	0.0664	.79730
4	.0706	.0545	1	80	1.3835	.24300
5	.1174	.1097	1	77	0.6730	.41455
6	.1174	.1144	1	77	0.2618	.61033
Category 9: Reaction to Combat Stress						
1	.1353	.1192	2	130	1.2078	.30219
2	.1192	.1183	1	132	0.1374	.71147
3	.1215	.1190	1	132	0.3819	.53768
4	.1190	.1180	1	133	0.1377	.71121
5	.1353	.1326	1	130	0.4063	.52497
6	.1353	.1212	1	130	2.1113	.14863
Category 10: Staying Ahead of the Situation and Keeping the A/C Informed During the Mission						
1	.1047	.0883	2	131	1.2046	.30311
2	.0883	.0774	1	133	1.5916	.20932
3	.0906	.0861	1	133	0.6517	.42094
4	.0861	.0753	1	134	1.5815	.21072
5	.1047	.0998	1	131	0.7245	.39623
6	.1047	.0801	1	131	3.6001	.05997
Category 11: Job Knowledge						
1	.1587	.1317	2	131	2.1006	.12648
2	.1317	.1291	1	133	0.3894	.53370
3	.1304	.1301	1	133	0.0507	.82218
4	.1301	.1275	1	134	0.4004	.52794
5	.1587	.1546	1	131	0.6274	.42973
6	.1587	.1508	1	131	1.2186	.27165
Category 12: Attitude Toward Job Performance						
1	.0722	.0415	2	131	2.1732	.11790
2	.0415	.0395	1	133	0.2729	.60225
3	.0378	.0327	1	133	0.7052	.40255
4	.0327	.0309	1	134	0.2569	.61306
5	.0722	.0719	1	131	0.0500	.82335
6	.0722	.0569	1	131	2.1634	.14373
Category 13: Overall Performance						
1	.1224	.0934	2	130	2.1471	.12095
2	.0934	.0910	1	132	0.3499	.55518
3	.0995	.0933	1	132	0.9172	.33996
4	.0933	.0909	1	133	0.3503	.55493
5	.1224	.1184	1	130	0.5902	.44375
6	.1224	.0948	1	130	4.0844	.04534

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Personnel Research Division Lackland Air Force Base, Texas 78236	2b. GROUP

3 REPORT TITLE

PROFICIENCY DIFFERENCES OF PILOT AND NAVIGATOR F-4 SECOND-SEAT CREWMEMBERS: ~~A~~SOUTHEAST ASIA EVALUATION

4 DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5 AUTHOR(S) *(First name, middle initial, last name)*

Shore, C. Wayne, Capt, USAF
Curran, Charles R., Capt, USAF
Ratliff, Forrest R., Lt Col, USAF
Chiorini, John R., Capt, USAF

6 REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
April 1970	19	4

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO 6323 Task c. 632305 d.	AFHRL-TR-70-9 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10 DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
	Personnel Research Division Lackland Air Force Base, Texas 78236

13 ABSTRACT

A proficiency rating form was administered to F-4 crewmembers following each of a series of combat missions. The rating form was one of several data collection instruments developed as part of Project Combat Team to gather behavioral data designed to relate the differences in the training of pilots and navigators to operational performance. The rating categories were designed to measure proficiency in ten second-seater functions and three general characteristics related to second-seater proficiency. Using an 11-point rating scale, aircraft commanders compared their second-seat crewmember with second-seaters of equal combat experience. Differences between pilot and navigator second-seaters on each rating category were tested by a multiple linear regression analysis. Data from other instruments relevant to evaluation of mission success were also presented. Several conclusions were drawn on the basis of these data: (*a*) It is possible to obtain reliable, valid proficiency data in a combat environment. (*b*) Proficiency ratings can be used to measure the performance differences between two groups in an operational setting. (*c*) By the 30th mission neither group of second-seaters demonstrated a general superiority in performance over the other. (Differences between pilots and navigators early in their combat experience were sharply reduced by mission segment 26 to 30, except in performance of the task Understanding Radio Communications.) (*d*) The early proficiency differences between pilot and navigator second-seaters appeared to be a function of training differences. Therefore, it may be possible to reduce or eliminate these differences through modifications in crew training programs. (*e*) During the first 30 combat missions, pilots flew on more missions rated as completely successful than did navigators. Although the differences were small, they were statistically significant.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Project Combat Team						
aircraft commander						
second-seat crewmember						
F-4 aircraft						
pilot						
navigator						
proficiency ratings						
mission effectiveness						
mission background data						
combat evaluation						